MSDS 413, Summer 2019, Assignment 2
Forecasting - Fundamental Concepts (TS2)

Introduction

For this assignment, you will use the datasets d-nflx3dx0913.txt and ukcars which was included in the Session 2 zip file (TS2.zip), and, along with a R script (TS2.R), is included on Canvas. You will read the file into R or other statistics package and conduct the requested analyses.

The following list defines the variables:

- d-nflx3dx0913.txt
  - PERMNO: CRSP 5-digit integer for all common securities
  - date: yearmonthday as, e.g., 20150131
  - nflx: Netflix company stock code
  - vwretd: CRSP value-weighted index
  - ewretd: CRSP equal-weighted index
  - sprtrn: S&P composite index
- Quarterly UK passenger vehicle production data (data set ukcars) from the Hyndeman text (in expsmooth via fpp packakge) as an array of
  - year (beginning with 1997 and ending with 2005)
  - quarter (1, 2, 3 , 4, in sequence by record)
  - production quantity

Your objective is to explore the relationship among the variables, identify what types of data each variable is, and respond to each part in the Procedure Section that follows.

Procedure

The following steps are necessary to complete this assignment. Address each and every part and ensure that you cover all the details specified in the questions.

1. **EDA** (2 points) Consider the daily simple returns of Netflix (NFLX) stock, CRSP value-weighted index (VW), CRSP equal-weighted index (EW), and the S&P composite index (SP) from January 2, 2009 to December 31, 2013. Returns of the three indices include dividends. The data are in the file d-nflx3dx0913.txt and the columns show PERMNO, date, nflx, vwretd, ewretd, and sprtn, respectively, with the last four columns showing the simple returns.

1.1. Conduct, contrast, and compare the EDA of the raw data and the log-transformed data of returns of NFLX stock and the S&P composite index.

1.2. Obtain the empirical density plot and QQ plot of the daily log returns of NFLX stock and the S&P composite index. Interpret the plots.

1.3. Test the null hypothesis that the mean of the log returns of NFLX stock is zero. Interpret the test outcome.

1.4. What are the pairwise associations between the log-transformed returns of NFLX stock and the S&P composite index? Why examine these associations?

2. **Hypothesis tests** (2 points) Consider again the daily simple returns of Netflix (NFLX) stock, CRSP value-weighted index (VW), CRSP equal-weighted index (EW), and the S&P composite index (SP) from January 2, 2009 to December 31, 2013. Returns of the three indices include dividends. The data are in the file d-nflx3dx0913.txt and the columns show PERMNO, date, nflx, vwretd, ewretd, and sprtn, respectively, with the last four columns showing the simple returns.

2.1. Test the null hypothesis that the log return ($\ln(nflx)$) is symmetric with respect to its mean. Interpret.

2.2. Test $H_0 : K = 0$ versus $H_a : K \neq 0$, where K denotes kurtosis. Interpret.

2.3. Construct a 95% confidence interval (CI) for the expected daily log returns of Netflix stock. Interpret.

2.4. Describe the meaning of the CI as if you are explaining it to a non-analyst stakeholder.

3. **Time series models** (2 points) use the quarterly UK passenger vehicle production data from 1997:1–2005:1 (data set ukcars) from the Hyndeman text.

3.1. Make a time plot of the data and describe the main features of the series.

3.2. Decompose the series using STL and obtain the seasonally adjusted data. Interpret.

3.3. Forecast the next two years of the series using an additive damped trend method applied to the seasonally adjusted data. Interpret.

3.4. Then re-seasonalize the forecasts. Record the parameters of the method and report the RMSE of the one-step forecasts from your method. Interpret.

3.5. Forecast the next two years of the series using Holt's linear method applied to the seasonally adjusted data. Then reseasonalize the forecasts. Record the parameters of the method and report the RMSE of the one-step forecasts from your method. Interpret.

3.6. Now use ets() to choose a seasonal model for the data. Justify your choice.

3.7. Compare the RMSE of the ets()-fitted model with the RMSE of the model you obtained using Holt's STL decomposition method. Which gives the better in-sample fits?

3.8. Compare the forecasts from the various approaches to justify your choice of which is best.

4. **Report** (1.5 points) Describe your choice of UK passenger vehicle model as if to a lay stakehoder. Which model would you use for forecasting, e.g.? The report requires information from which the stakeholder can make decisions or take action, and not encumbered with stat jargon.

.

Deliverables

See Section Submission Directions below. The assignment deliverables, each in pdf format, are as follows:

- *Only if requested by instructor*
  - The program or script
  - Logs
  - Outputs
- Mandatory
  Data analysis write-up: no programs, logs, or just code outputs.

The data analysis must follow and use the item numbering of each assignment, i.e., use the numbers, say, 1 - 5, with the sub-lettering if used. These deliverables are provided according to the instructions in the Submission Directions section below.

Submission Directions

Title Page

Include a title page with your name and the assignment designation. Leave room for instructor comments.

File Names

The assignment write-up file shall be submitted to Canvas according to the schedule in the syllabus using the item (1) naming convention below. The naming convention is case sensitive. Use letters and numbers as given. **The file name parts have no spaces or other separator charancters**.
TS2Lastname.pdf (submit via Canvas)

The parts are the assignment code, TS2; your lastname with only the first letter capitalized; a period, and lastly, the extension "pdf". Generically,

TS2Lastname.pdf

For example: Suppose your name is Student McStats. Your filename then is:

TS2Mcstats.pdf

The analysis write-up file must be submitted for grading. Each write-up requires a title page for instructor comments. The analysis may use either R or any other statistics package you wish, or if you use more than one package, you must use the germane tables, plots, etc., in a single report. If you use more than one package, differences and similarities should be indicated.

email: jamie.riggs@northwestern.edu

Email *ONLY IF REQUESTED* the program (script), log and output as separate pdf files. The R log and output may be combined. The file names shall be as follows:

- The program or script file names
  - TS2LastnameRprog.pdf
- The log file names
  - TS2LastnameRlog.pdf